

Verification of FE-D Digital Readout Performance

K. Einsweiler, LBNL

Overview of DMILL corner models used

Description of simulation strategy

Detailed analysis of typical case

Highlights of corner analysis, and weak aspects in design

SPICE Simulations of DMILL

TEMIC provides standard pre-rad corner models (ss, fs, sf, tt, ff). Models should cover process variations. Some of the critical parameters are monitored during fabrication:

- Threshold (V_{T0}) is varied: nominal NMOS = 750mV, PMOS = -800mV. The corners have +/-100mV variations in this parameter.
- Other parameters are varied: gate oxide thickness $TOX = 17.5\text{nm}$, corners are about +/-5%. The mobility U_0 is varied by about +/-5%. The effective W and L are varied, $LINT$ by +/-150nm, $WINT$ by +/-80nm. Some device capacitances are also varied from nominal by +/-5%. These should cover fab tolerances.

TEMIC, in collaboration with CPPM, has provided a set of BSIM3 models intended for corner analysis after 25MRad:

- The only post-rad parameter variations are in V_{T0} and U_0 , corresponding to V_t shifts and g_m reduction in devices.
- Worst case are used, with NMOS rebound and large PMOS shifts to reflect possible post-anneal (low dose rate) behavior.
- NMOS V_{T0} ranges from -200mV (iff) to +300mV (iss), with U_0 reduction of 25% for all post-rad case.
- PMOS V_{T0} ranges from -100mV (iff) to -400mV (iss), with U_0 reduction of 15%.

Results from TEMIC for FE-D1 runs:

- First run (FE-D1) had VTN08 of about 900mV, and VTP08 of about 800mV. The only parameter somewhat outside of the nominal tolerances was BVCEOT, which was slightly below the nominal 5.5V window on a number of PM structures.
- First run had rad-hard qualification on one wafer. The initial VTN08 was 960mV, GMN08 was 0.30mS, VTP08 was -860mV, and GMP08 was 0.13mS. The post-rad values were VTN08 of 830mV and GMN08 0.26mS, VTP08 of -1070mV and GMP08 of 0.11mS.
- Cannot compare measured values directly with model parameter, but V_t shifts of 160mV (NMOS) and 210mV (PMOS), and gm reductions of about 15% are certainly covered by the corner models we have.
- Second run (FE-D1b) had VTN08 of 870mV, and VTP08 of -850mV. All measurements were well-centered in the windows.

Conclude:

- From parametric point of view, both runs fairly typical, with no large differences between them.
- Although the parametric changes after 10MRad doses are inside the corner models, they are larger than has been advertised (typically $\Delta V_{th}=100\text{mV}$).

Simulation Strategy

Environmental conditions:

- Power budget and specifications are written in terms of a “typical” and “worst case” set of power supply voltages. The typical values are taken to be 3.0V/1.5V VDDA/VCCA (analog supplies) and 3.0V VDD (digital supply). The worst case values are taken to be 3.5V/1.75V for VDDA/VCCA and 4.0V for VDD.
- The range of operating temperatures for the chips is tightly constrained by the module design, mechanics and cooling. The chips themselves will be in direct contact with the cooling structure, and so their temperature should be in the range of -30C up to 0C (typical operating temperature should be in range of -7C to 0C). We do all simulations at room temperature as a worst case. More work should be done on low-temperature device models and simulations for the future.

Design Goals:

- We are designing for operation of the chips at nominal voltage pre-rad (for all corners) and worst-case voltage post-rad (for all corners). Ideally, we would have some extra margin in both cases (say XCK OK up to 50MHz).
- One method of assessing the margin in the design is to increase the master clock frequency (XCK) beyond the 40MHz nominal value. We plan on using this approach extensively during production testing.
- Have observed in several different simulations that 3V tt is about twice as fast as 4V iss. Therefore, we have looked at operation up to 80MHz pre-rad and 50MHz post-rad.

Simulation method:

- We use Verilog for much of the logic verification in the design, but the approach used by Bonn for FE-D does not include back-annotation for capacitive loading due to routing, and hence cannot be considered a real timing simulation.
- We have tried to make the largest SPICE simulations possible, including all parasitics. This is unfortunately rather painful. It requires special “chopped” layouts to eliminate all of the circuitry that one doesn’t want to simulate. It also requires making corresponding schematics to run LVS and make sure that no errors have been introduced into the chopped layout.
- Running a typical simulation of this size is painful. We use ELDO-XL for this, which incorporates automatic generation of special “linearized” transistor models to accelerate the simulation (risky for designs with “analog” features). This gives us a typical speed-up of 5 over standard SPICE, and allows us to simulate a 50K transistor flat netlist with parasitics at a rate of about 5 μ s/day on a fast HP.

Standard “single column pair” simulations:

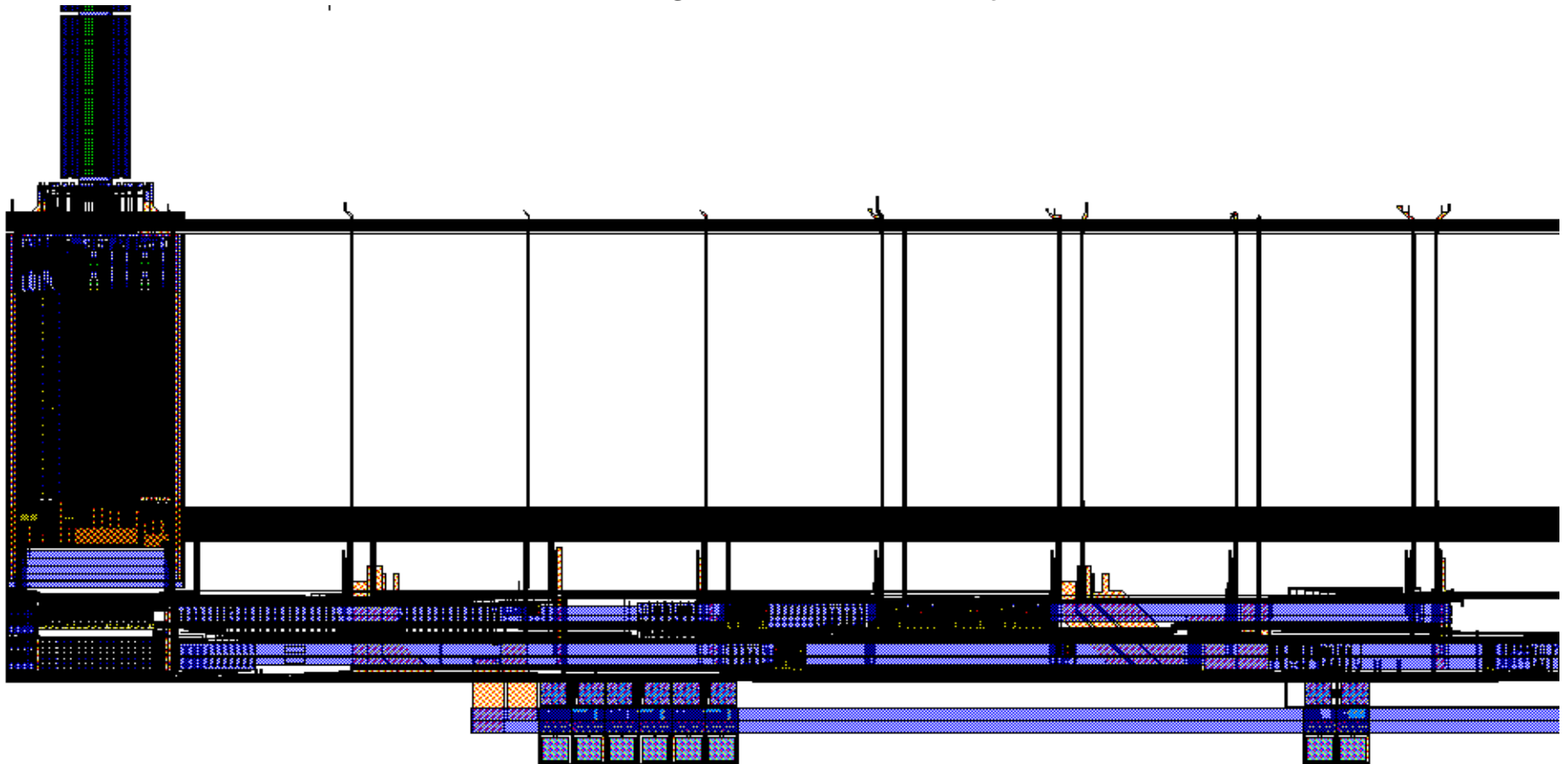
- A special chopped layout was constructed which included the digital back-end of the 320 pixels in a column pair, the complete EOC, and all of the digital logic required for readout at the bottom of the chip. All of the critical LVDS drivers and receivers are included, as well as the large internal metal busses used for routing of signals.

- The main issues that cannot be addressed by this netlist are those involving collection of the data from the nine column pairs in the chip. Fortunately, this is a very limited part of the design, with significant buffering for inputs and outputs at the bottom of each EOC block. The critical function to be simulated is the 2D sparse scan to transfer data from the EOC to the TOT subtractor and serializer.
- These remaining issues can be addressed using a nine column pair chopped layout, in which the minimum block of 16 pixels is left on top of each EOC block. These simulations have not been pursued at this time, because the timing of the signals measured in the single column pair case suggests that there is plenty of margin in the 2D sparse scan circuitry (and the netlist is much larger, close to 100K transistors).

Different designs simulated:

- Many simulations were performed on a chopped layout for the present FE-D1 design (this was done only after submission). These simulations show results consistent with those found in the lab when measuring chips.
- Many simulations have been performed of the improved FE-D2, in which the design of the dynamic logic in the pixel has not been changed (FE-D2D).
- A small number of simulations have been performed of the FE-D2 version in which the hit logic has been made fully static (FE-D2S).

- Some details of the chopped single column pair layout:

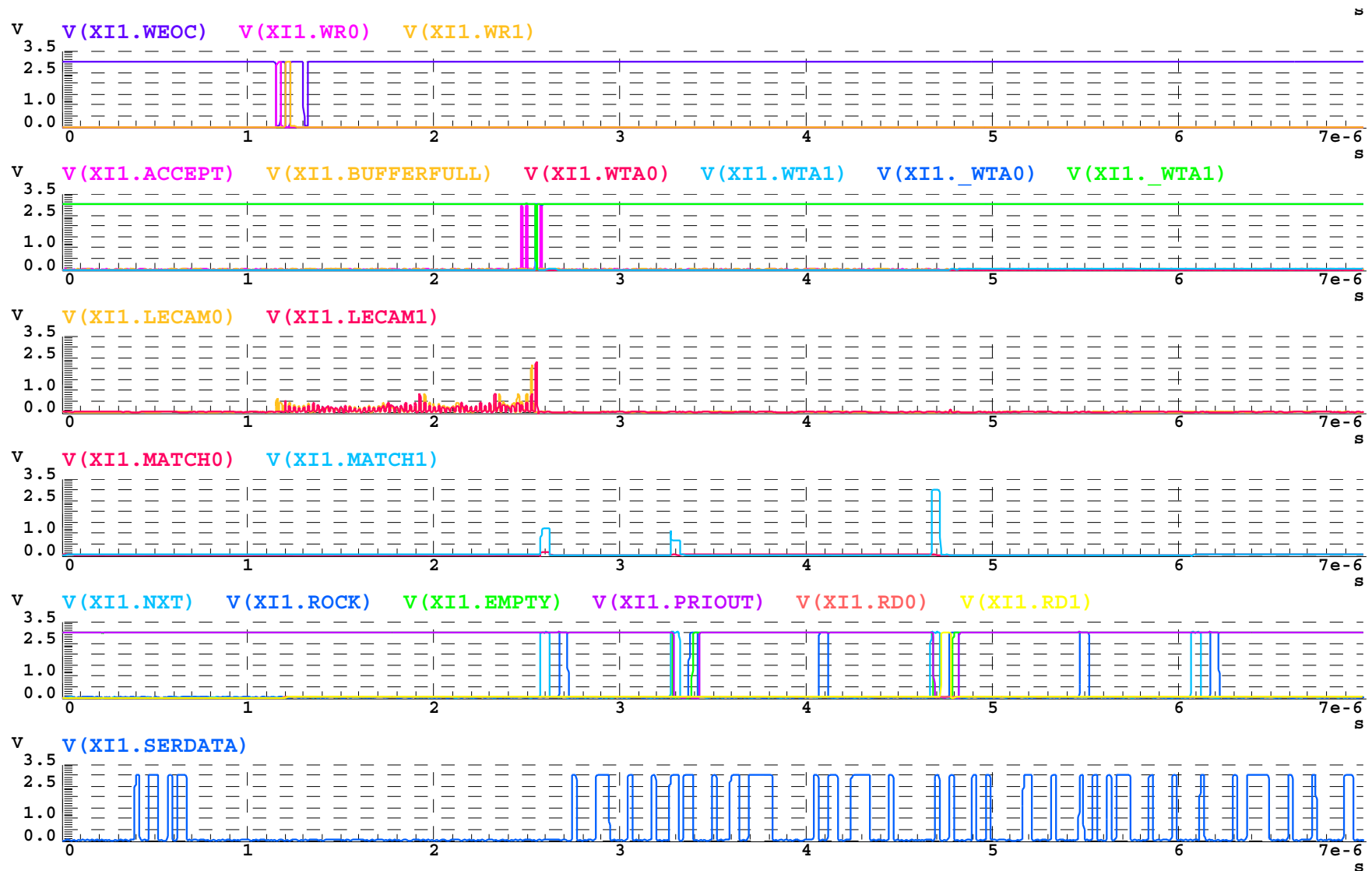


- Three LVDS inputs (XCK, LV1, SYNC) plus one LVDS output (DO) are present.
- All bottom digital circuitry is present (Grey generator, Clock generator, FIFO, RCU, TOT subtractor, Serializer, Output MUX, short Global Register).
- Column pair with 320 pixels and 24 EOC buffers is present.
- All routing (including vertical channels) and bussing of critical signals is present.

Simulation Results for Typical Case

- Extracted netlist has about 51K transistors and 83K capacitors.
- Typical test vector is used in which three hits were injected (two right, one left). The left hit was at the top of the column pair, the right hits were in the middle and near the bottom.
- Two groups of two triggers each were provided, separated by one empty crossing. One hit was associated with the second trigger in the first group, one was associated with the un-triggered crossing and one was associated with the first trigger in the second group.
- The relevant piece-wise linear SPICE inputs were generated from a Verilog description. This allows testing of the vectors with Verilog, and relatively painless generation of complex waveforms.
- This simulation should produce four EOE words, where the second and third include a single hit, and the other two are empty events.
- Note: Additional simulations were performed with a very large number of hits, just to check whether there are logical flaws in the readout design. No significant problems were uncovered. Most of these simulations were performed using a large netlist extracted from the schematic level prior to submitting FE-D1, rather than using the latest layout level simulation.

•Overview of test vector simulation:



- Hits enter EOC at about 1200ns. LVL1 trigger accepts hits at about 2500ns.
- Readout begins, followed by 4 NXT (events), and six data words.

Critical aspects of the design:

Transfer of hits from pixels to the bottom of the column.

- The hit readout procedure should operate at 20MHz (transfer one hit every 50ns).
- This requires a sparse scan procedure which first notifies the CEU when a hit is ready for transfer, and also autonomously selects the topmost unread hit for transfer in each column of the pair. This sparse scan operates as 10 16-pixel scans in parallel for speed, and it must complete in slightly less than 50ns.
- Writing to the EOC buffer block must also occur at 20MHz. It involves finding the next free EOC buffer (this is overlapped with the write of the previous hit), and storing the input data.

Making trigger accept/reject decision for each buffered hit.

- This procedure must always operate at 40MHz with no false comparisons.
- Once a hit is stored in a given EOC buffer, a 7-bit comparator is activated to compare the timestamp of the hit with the reference timestamp. If they agree on a crossing with a LVL1 trigger, the buffered hit is flagged as triggered. Otherwise, the buffer is reset and available for new data.
- To save space, this comparator use precharged dynamic logic, with 12.5ns to precharge and 12.5ns to evaluate.

Finding a free EOC buffer for writing hits, and finding hits in EOC buffers and reading them out.

- There are two independent, and conceptually identical, sparse scans involved in these operations. Both scans have a similar 50ns maximum time constraint.
- Whenever a hit is written, an overlapping scan is done to find the next free buffer for writing. This allows a new hit to be stored immediately when WEOC is set. It is essential that a new buffer always be selected when WEOC is issued, otherwise data will be lost. The worst case is hard to test in simulation.
- The hit readout involves a 2D sparse scan, with each EOC buffer block scanning downwards to find the first hit in the block. In parallel, a horizontal scan selects the first EOC buffer block with data.
- If this 2D scan does not complete in time, then the RCU will transmit an EOE word to signal there is no more data to transmit for the current trigger.
- The readout logic for data transfer in the bottom of the chip operates on a 10MHz clock (ROCK), where the first 50ns is available for the vertical EOC buffer scan, and the second 50ns is available for the horizontal scan through column pairs.

Procedure:

- The timing for each of these critical operations has been studied in detail for different XCK frequencies, different supply voltages, and different SPICE models.

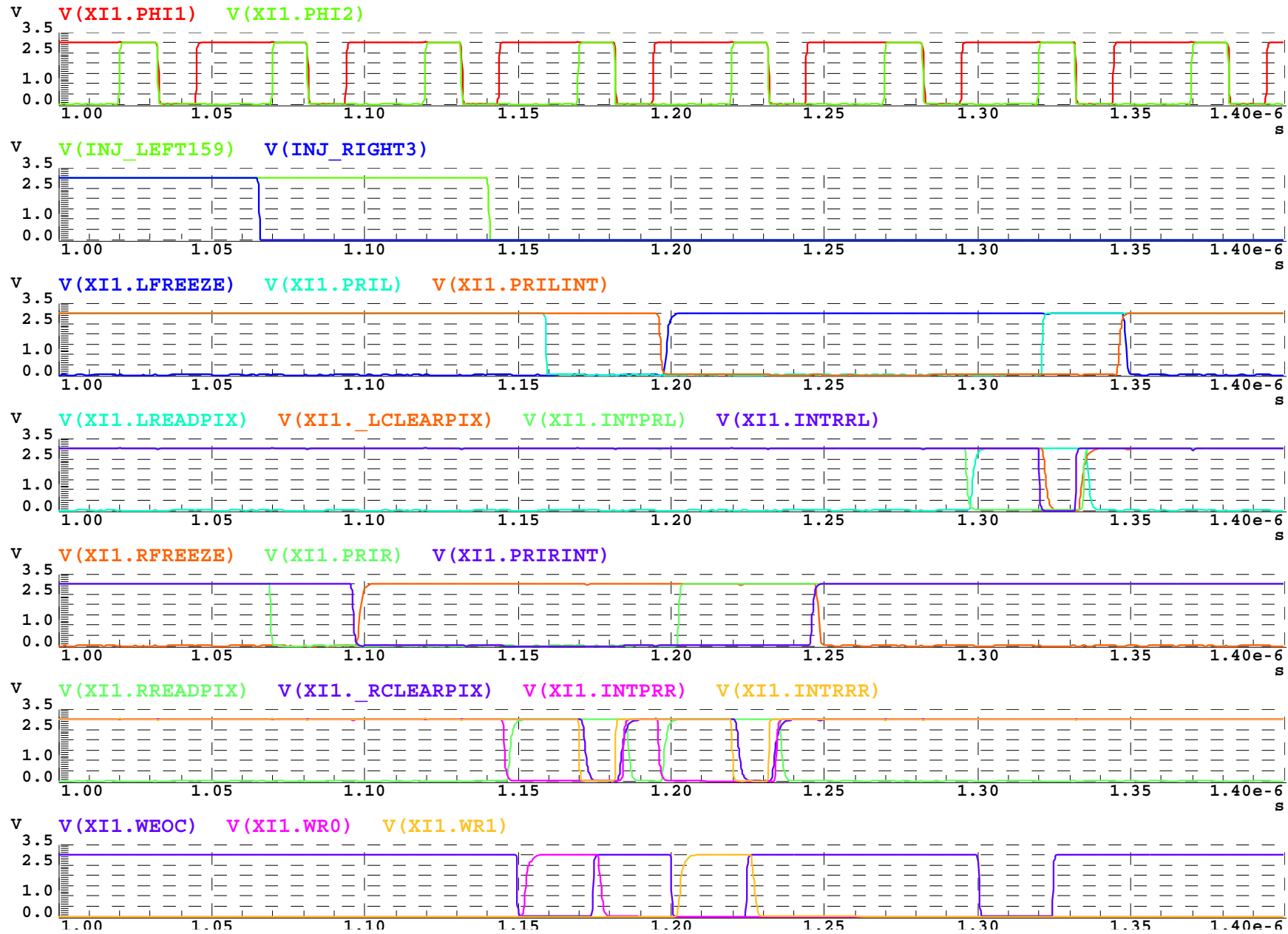
Hit Transfer Sequence:

- Trailing edge of hit initiates Pri scan. A low signal ripples down the column to the CEU. After a settling time, only the topmost hit pixel will be selected by the sparse scan circuitry (however, before settling, other values are possible).
- Once Pri reaches CEU, a Freeze is generated on the next Phi1 edge to prevent new hits from entering the sparse scan circuitry.
- The following Phi1 edge, a ReadPix signal is generated (37.5ns wide), and a ClearPix signal (12.5ns wide) is generated. Upon receipt of the ReadPix signal, the presently selected pixel releases the scan so it can select the next pixel in column to read out. The ClearPix signal clears the hit logic for the selected pixel.
- Each ReadPix signal generates a corresponding WEOC signal

Some critical times are:

- The time required for a hit at the top of the column to ripple down. It is important for the sparse scan to have adequate time to reach its “equilibrium” position to avoid selecting more than one pixel. For the Left159 hit, it takes about 20ns for the information to reach the CEU.
- The time required after a single hit at the top of the column has been read out, for the CEU to realize that there is no longer any data. If this is too slow, then the CEU will generate a spurious extra read cycle. In this case, it takes about 25ns for the Pri clearing operation to complete.

•Hit transfer simulation of FE-D2D with XCK=40MHz, tt parameters, VDD=3V:



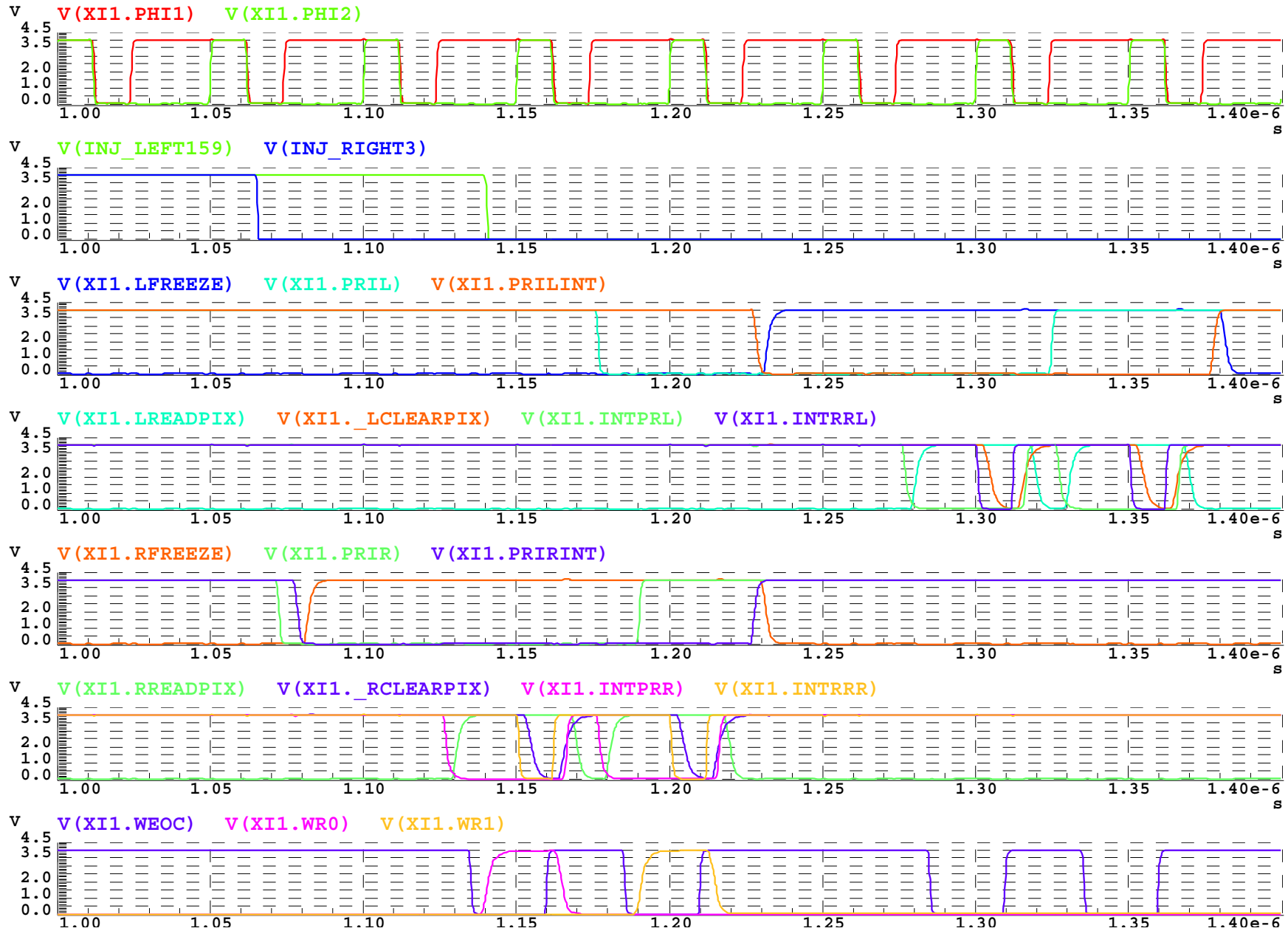
Summary of Pri Scan results for different corners:

- For ff at 3V, see about 14ns one direction, and 17ns the other way.
- For tt at 3V, see about 20ns one direction, and 25ns the other way.
- For ss at 3V, see about 31ns one direction, and 36ns the other way.
- For iss at 4V, scan needs about 38ns in one direction and 44ns in the other.

Conclusions:

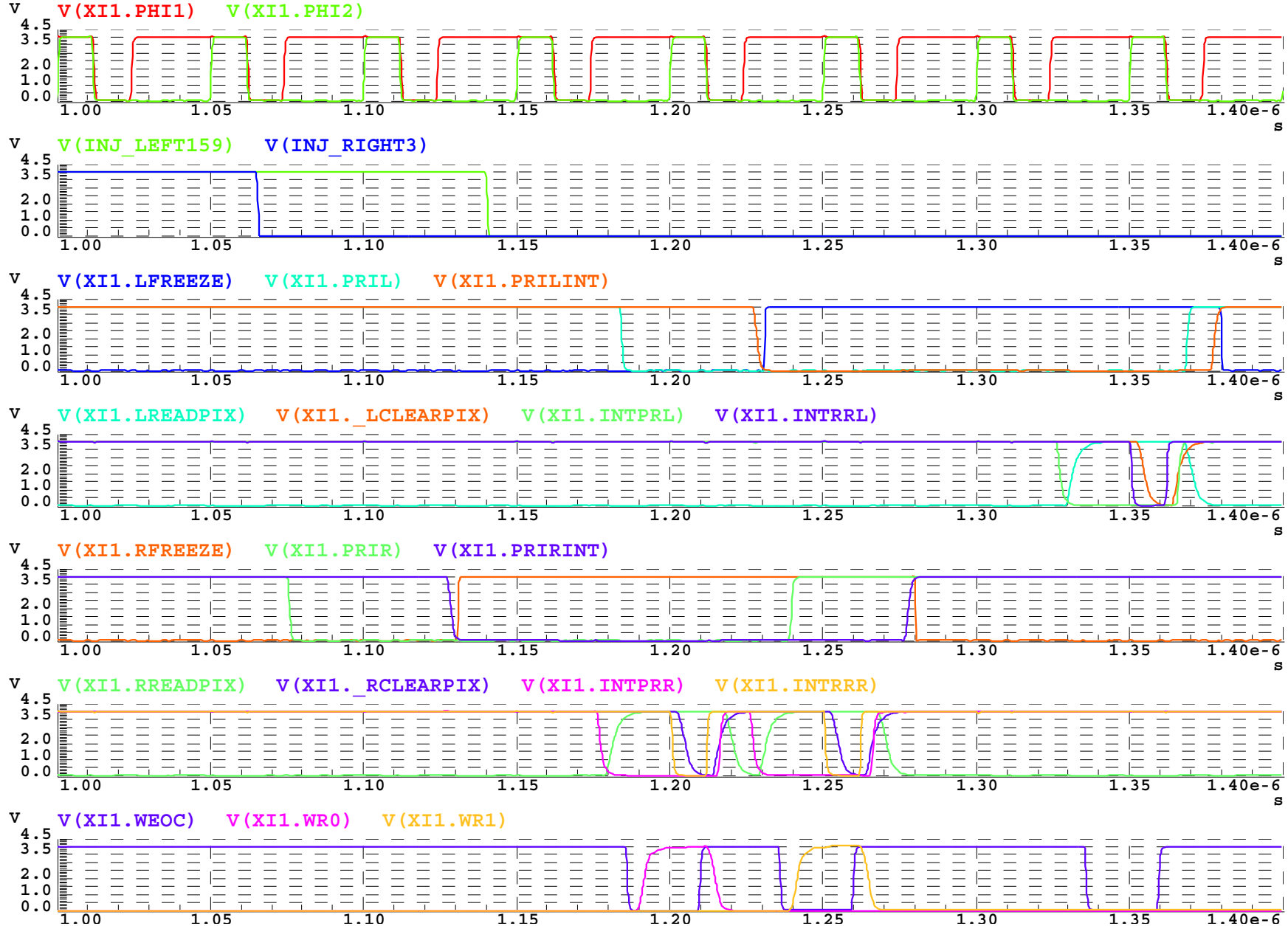
- It turns out that instead of the nominal 50ns, something closer to 45ns the time available at 20MHz readout rate. This is for the worst case where there is a single pixel at the top of the column pair, and the CEU must suppress a second ReadPix after this hit is read, and is due to small overheads in the CEU logic.
- It is difficult to do much better, as increasing drive strength in the sparse scan chain increases capacitance, and to first order, these effects compensate. However, the sizing in FE-D2S is probably more optimal than in FE-D2D.
- The column sparse scan operates up to a XCK frequency of 80MHz with tt models and VDD=4V.
- The column sparse scan just barely fails to work properly at XCK=40MHz and iss models and VDD=4V. This can be slightly improved by transistor sizing, but significant re-layout is required to squeeze in the improved sizes.

•Hit transfer simulation of FE-D2D with XCK=40MHz, iss parameters, VDD=4V:



- In this case, the release of Pri after the first ReadPix in the Left column takes too long. The CEU goes ahead and generates a second ReadPix (which has no corresponding hit), which will generate a spurious hit with typically zero values in all of the data fields.
- This occurs in part because the relative sizing of NMOS and PMOS in the NAND2 and NOR2 of the sparse scan is not optimal (the NMOS needs to be larger in NAND and the PMOS needs to be larger in NOR). These changes were performed in the FE-D2S version of the layout.

•Hit transfer simulation of FE-D2S with XCK=40MHz, iss parameters, VDD=4V:

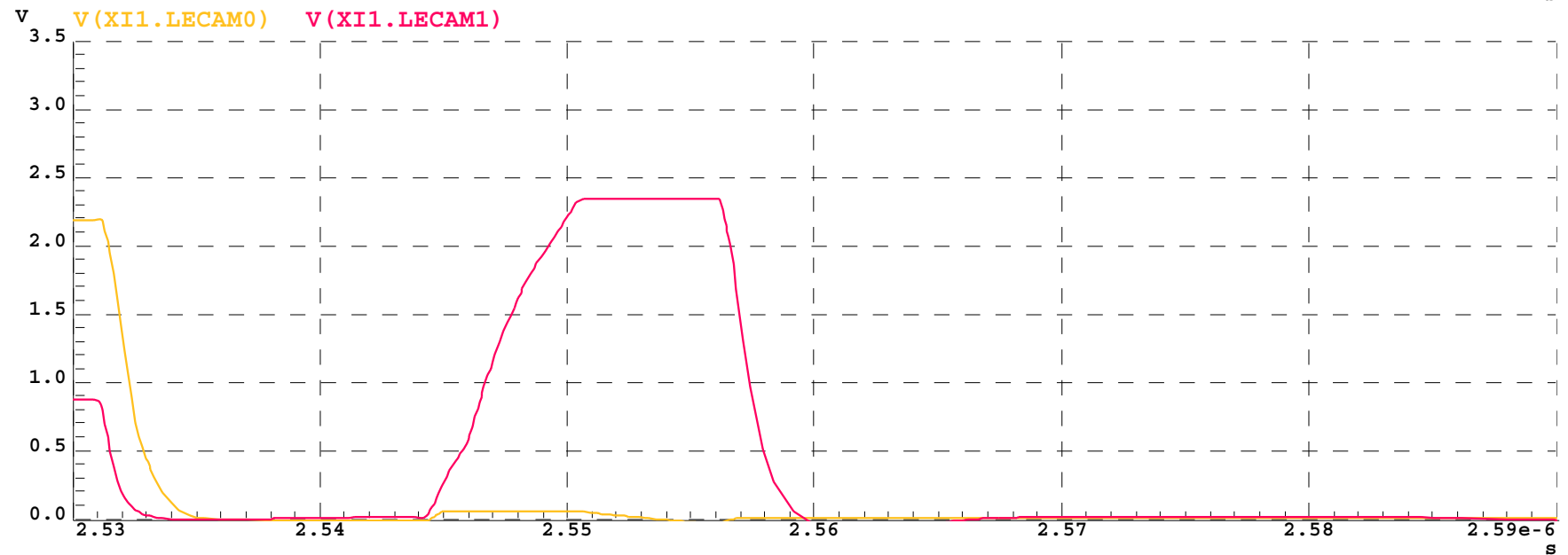
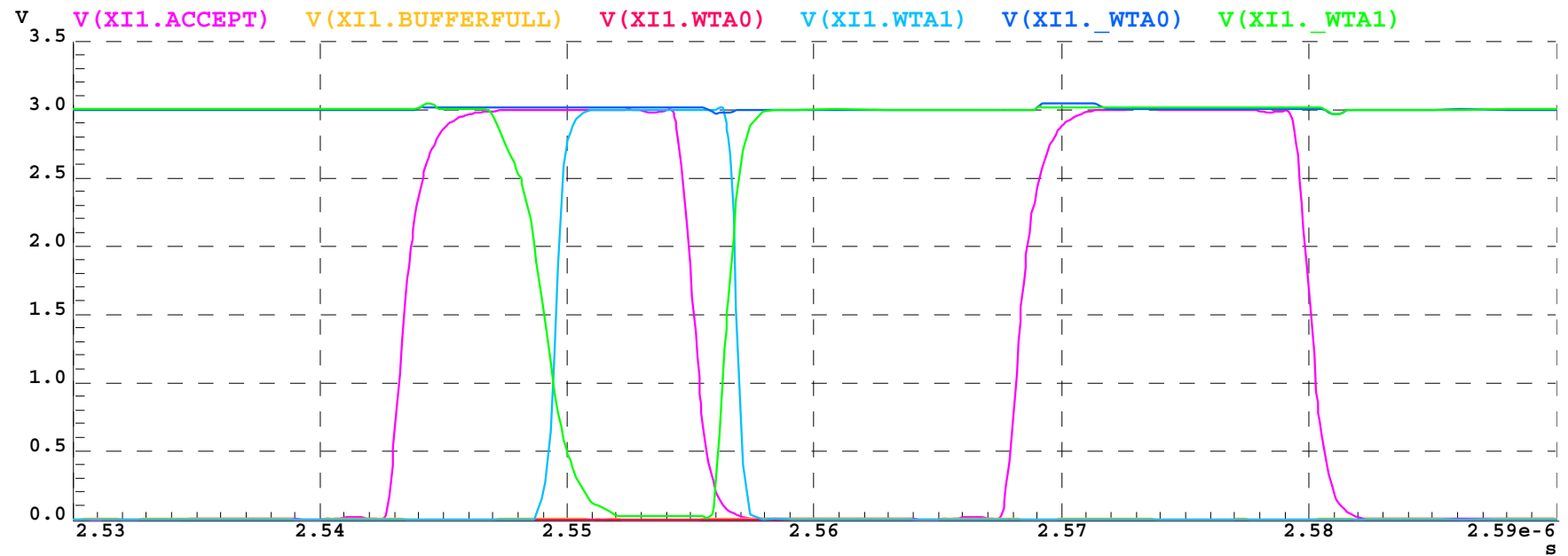


- First observation is that the modified netlist caused ELDO to start the ClockGenerator for Phi1 and Phi2 with a different phase. This changes the order of the hit readout compared to all of the previous FE-D2D simulations.
- The modified NAND and NOR sizings degrade the initial scan time (now about 43ns instead of 38ns) for Pri to go low. They have the opposite effect on the Pri high transition, which is sped up from 44ns to 35ns. This is enough to eliminate the spurious hit produced before. It also indicates the difficulty of optimizing sizing, due to the drive strength versus capacitive load trade-offs.

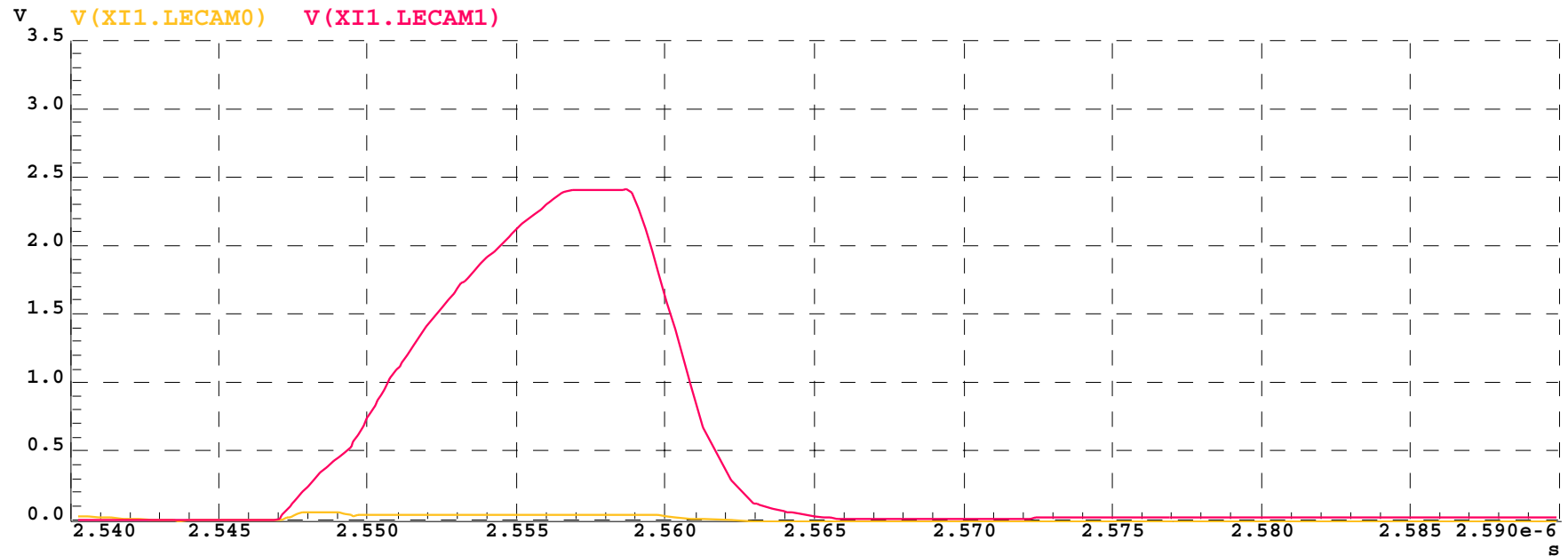
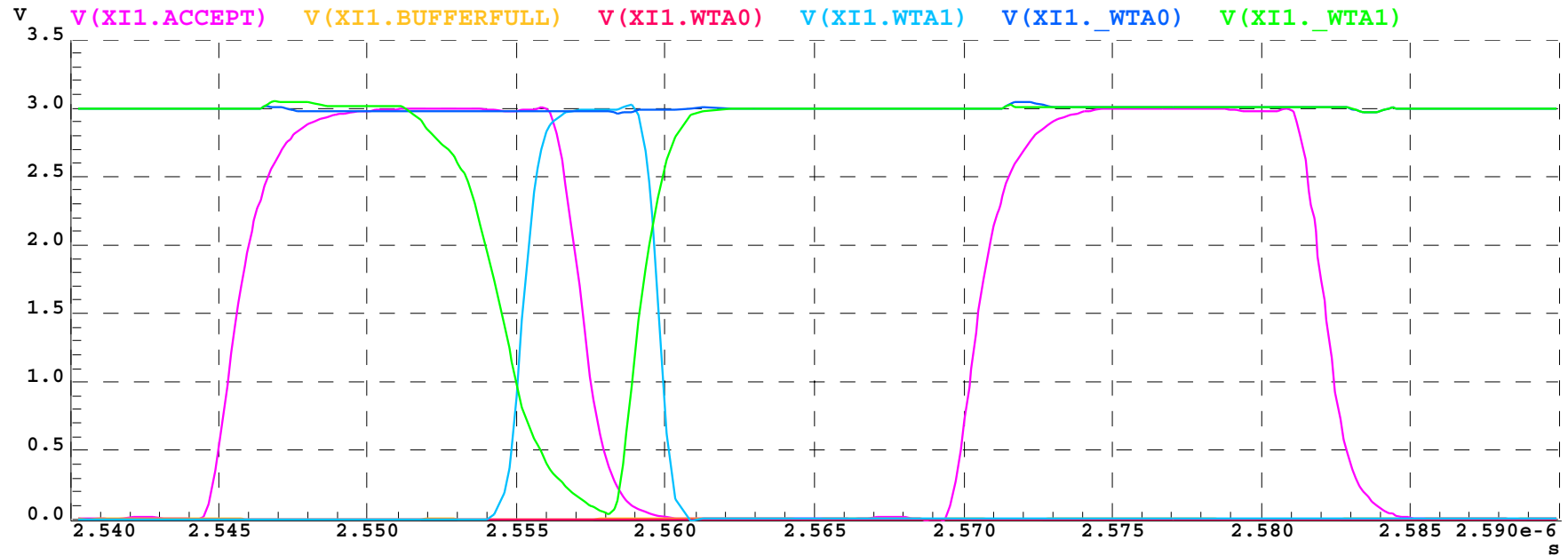
Trigger Coincidence Sequence

- Once a buffer is flagged as occupied, the 7-bit CAM will operate every 25ns to check the stored LE timestamp against the external TSC timestamp.
- The CAM must have proper transistor sizing so that it produces a reliable match signal when all bits agree, and a signal which is clearly below threshold for the single bit mis-match case.
- The CAM is precharged in the first 12.5ns, and then evaluates during the next 12.5ns. If the CAM match signal and a trigger accept overlap, then a WTA signal is generated to store the present trigger number into the 4-bit CAM in the EOC buffer, and to set the ValidTrigger FF.
- If the CAM match signal and a reject (absence of accept) occurs, then the EOC buffer occupied status is cleared.
- If the CAM is a little slow, then the WTA does not occur, but the CAM match is wide enough to reset the occupied state, and the hit is lost. If the CAM is even slower, then even the reset cannot occur, and buffers will never be cleared until a hardware reset is issued.

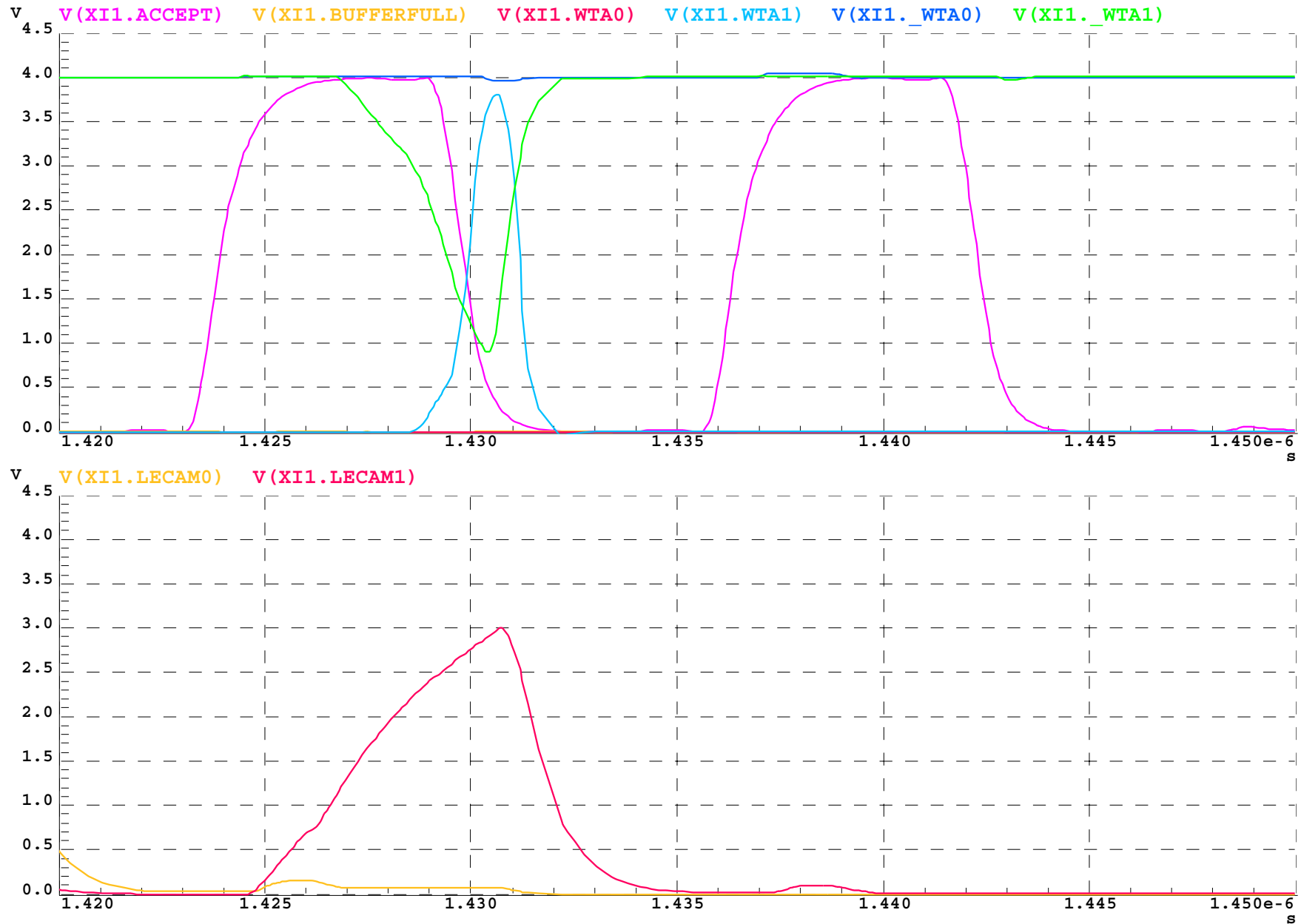
•Trigger match simulation of FE-D2D with XCK=40MHz, tt parameters, VDD=3V:



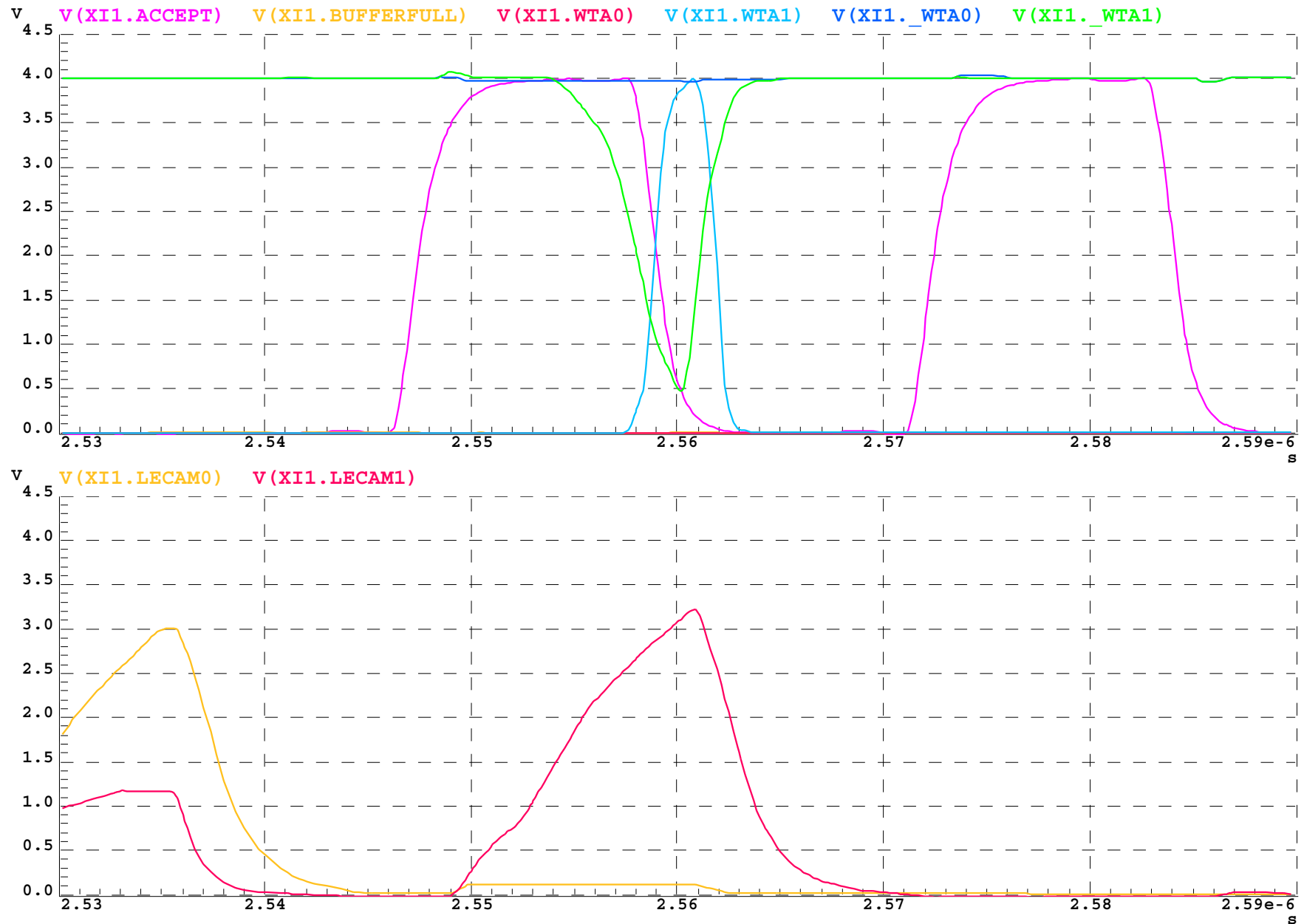
• Trigger match simulation of FE-D2D with XCK=40MHz, ss parameters, VDD=3V:



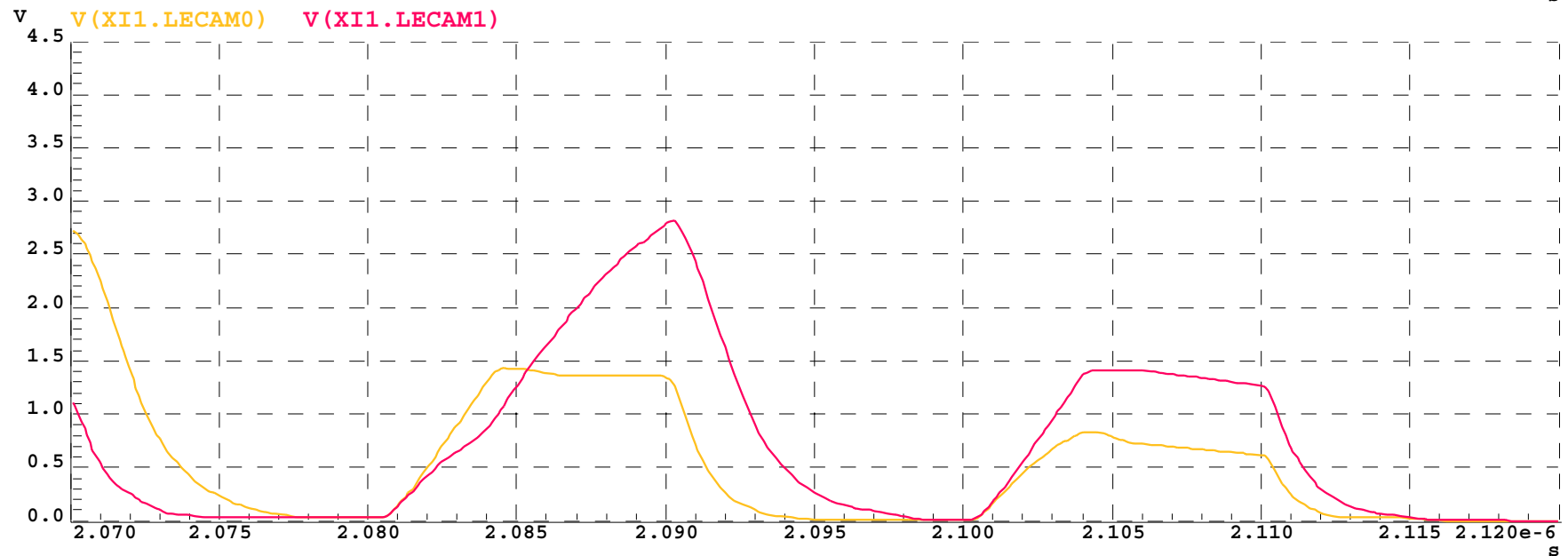
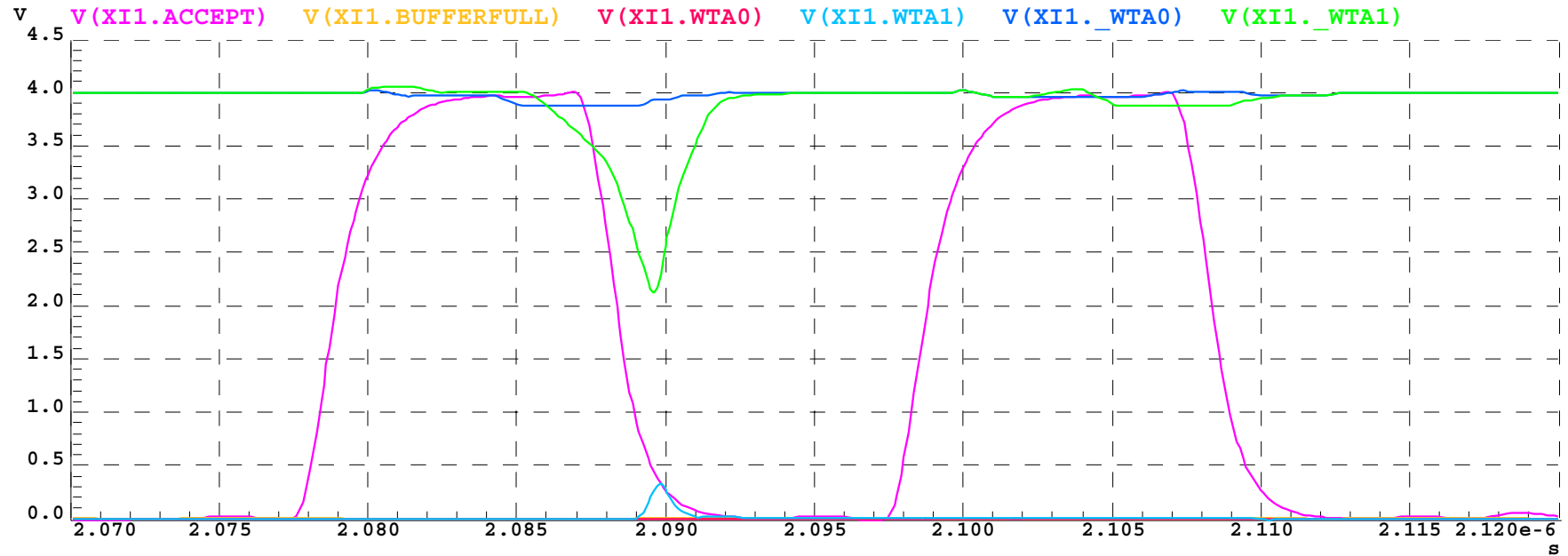
• Trigger match simulation of FE-D2D with XCK=80MHz, tt parameters, VDD=4V:



• Trigger match simulation of FE-D2D with XCK=40MHz, iss parameters, VDD=4V:



• Trigger match simulation of FE-D2D with XCK=50MHz, iss parameters, VDD=4V:



Summary of WTA Results for different corners

- For ff at 3V, see WTA pulse of 8ns, with CAM evaluation time of 4ns.
- For tt at 3V, see WTA pulse of 7ns, with CAM evaluation time of 6ns.
- For tt at 4V and 80MHz, see WTA pulse of 1ns, with incomplete CAM evaluation.
- For ss at 3V, see WTA pulse of 4ns with CAM evaluation time of 10ns.
- For iss at 4V, see WTA pulse of 3ns, with CAM evaluation time of more than 12ns.
- For iss at 4V and 50MHz, see no WTA pulse, with incomplete CAM evaluation.

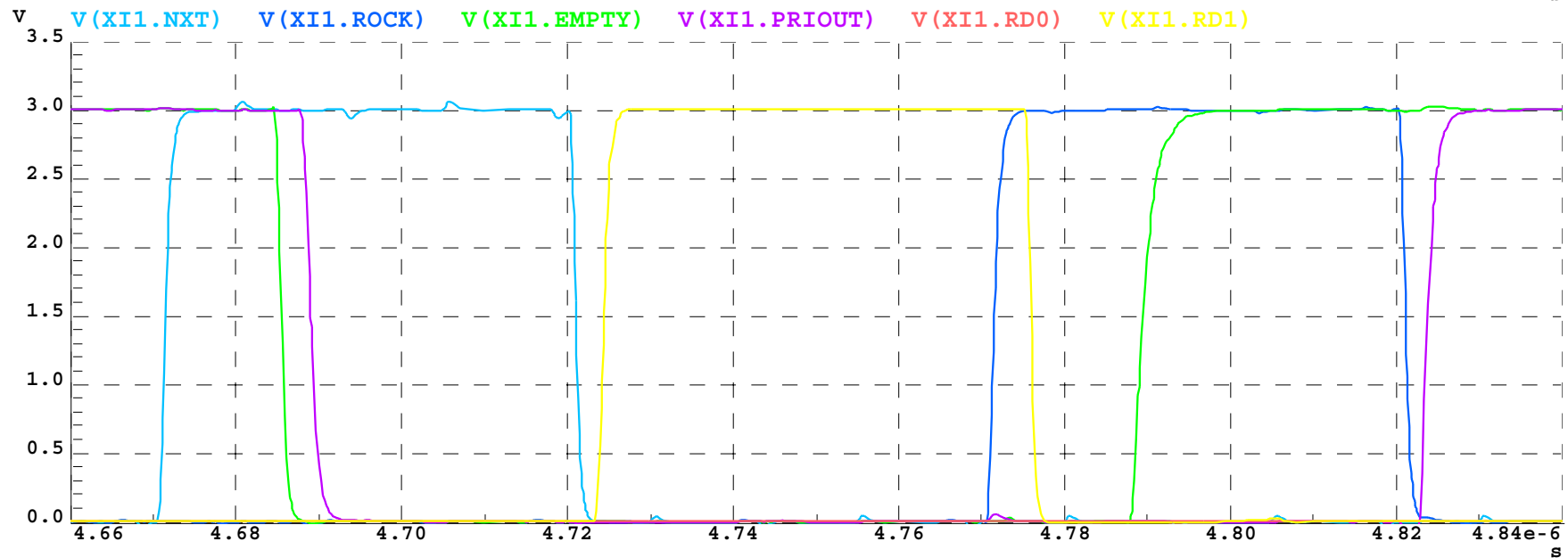
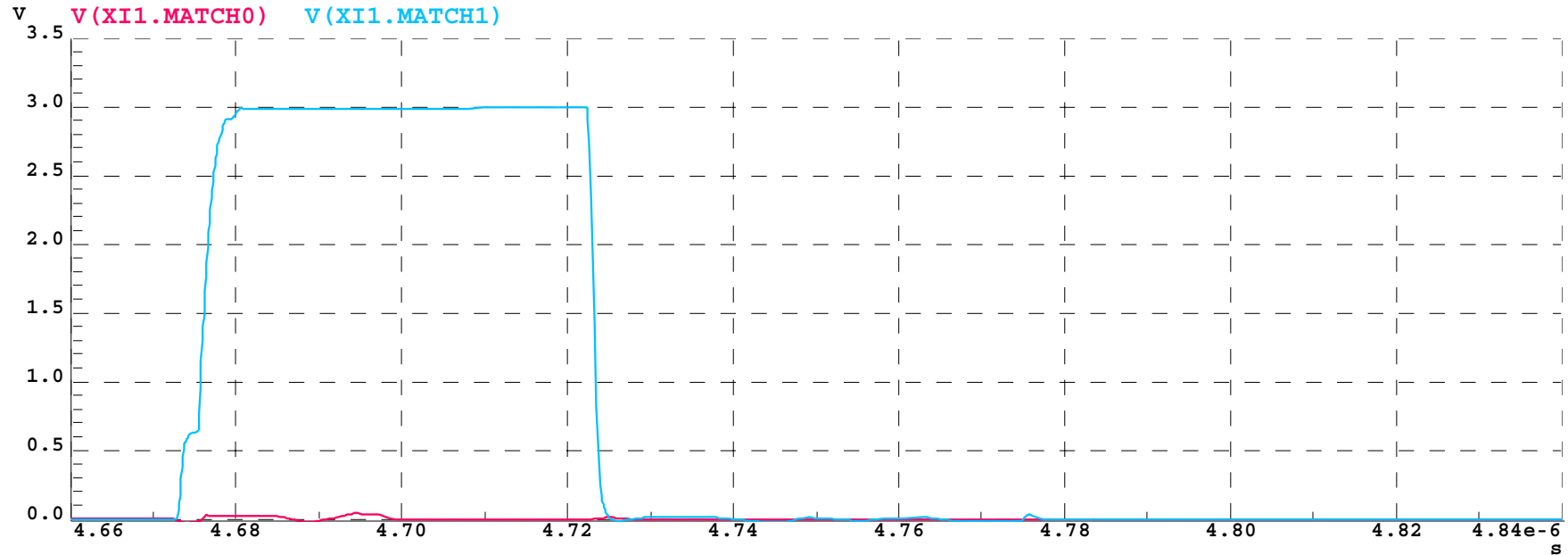
Conclusions:

- For both the tt 80MHz simulation at 4V and the iss 40MHz simulation at 4V, the write is very marginal. In both cases, the simulation works, and the chip produces the correct result. However, it is clear that this is very close to the edge. In fact, the iss simulation at 50MHz fails, and all EOC buffer entries are cleared (the CAM output is just wide enough to clear but not wide enough to accept).
- In looking at the waveforms and the schematic, it is clear that there is a drive strength problem with the NAND which makes the coincidence with the trigger accept (and generates the green _WTA1 signal). This gate has two 2.2 μ NMOS in series (effectively 1/2 min size), and drives a pair of inverters, and the four RAM bits in the TA CAM. From the waveforms, it is clear that several ns are lost due to the poor high->low transition time on this signal.

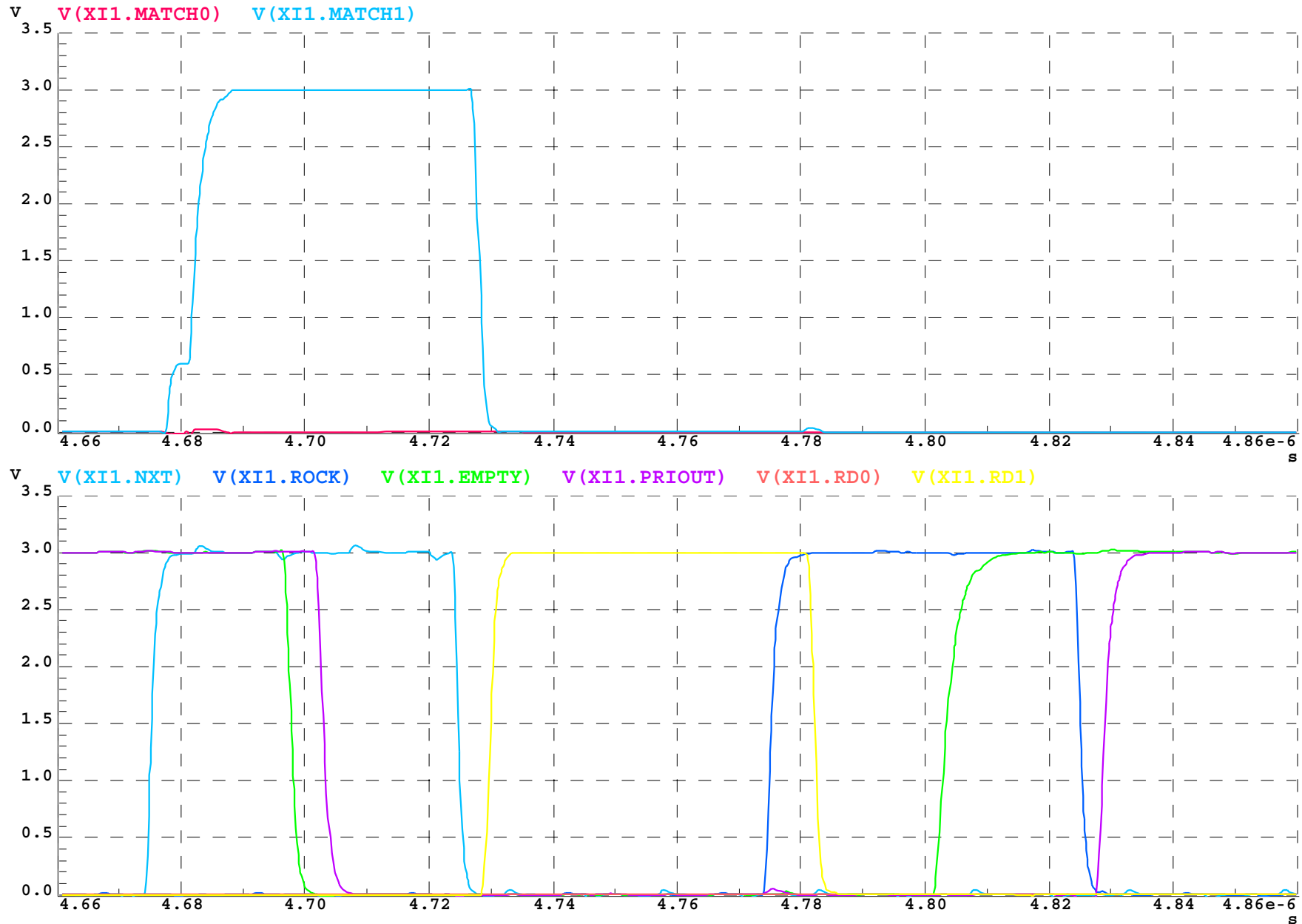
2D Sparse Scan Sequence

- Scan is initiated by NXT, which starts the 4-bit TA CAM evaluation sequence.
- The results of this evaluation are entered into a sparse scan tree just like in the column. This sparse scan will select the topmost EOC buffer which has a matching TA value. The sparse scan propagates to the bottom of the EOC buffer block and emerges as the EMPTY signal, with low indicating there is at least one matching hit.
- In the present simulation, the hit which matches is in the second of the 24 buffers in the block, and so the time from the NXT transition to the EMPTY transition is a measure of the total sparse scan time in the EOC block.
- Once the NXT goes away, the hit information is presented to the bus across the bottom of the EOC blocks (RD signal turns on), and the data is latched on the leading edge of the readout clock (ROCK) into the serializer.
- The ROCK transition also resets the EOC buffer and lets the sparse scan continue propagating, so the time until the EMPTY transition is a measure of the sparse scan speed in the other direction.
- Finally, the transition on PRIOUT indicates all data has been transferred, and will initiate the transmission of the EOE word.

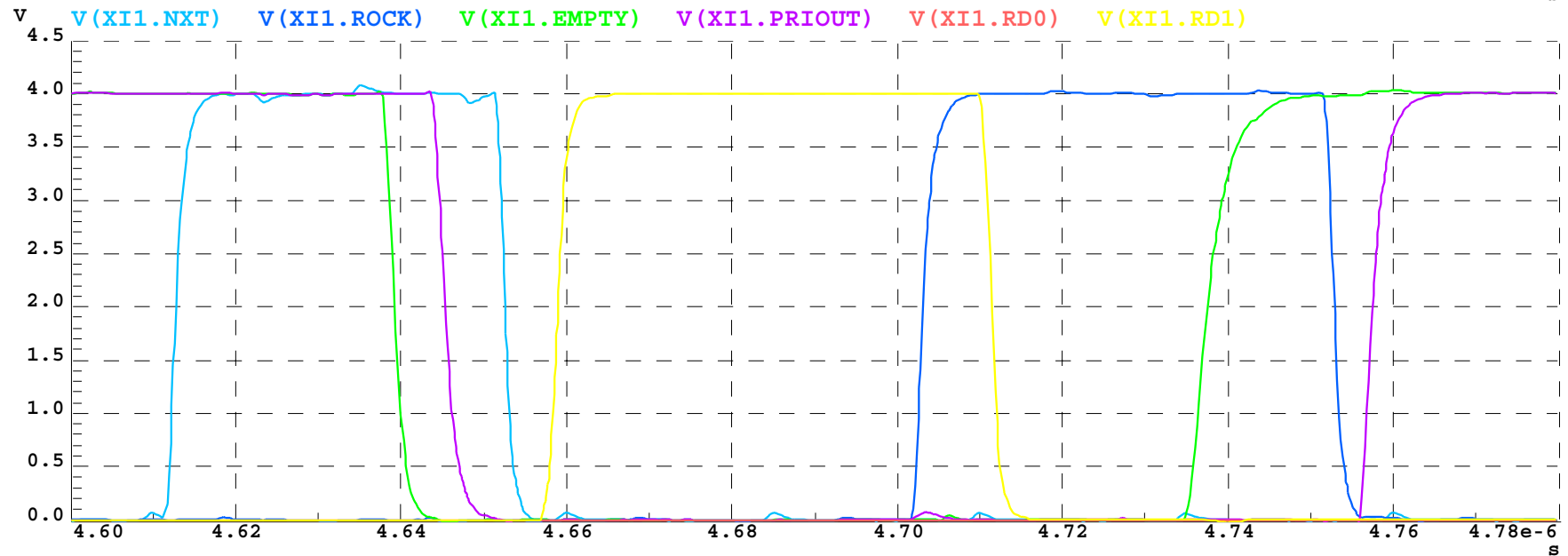
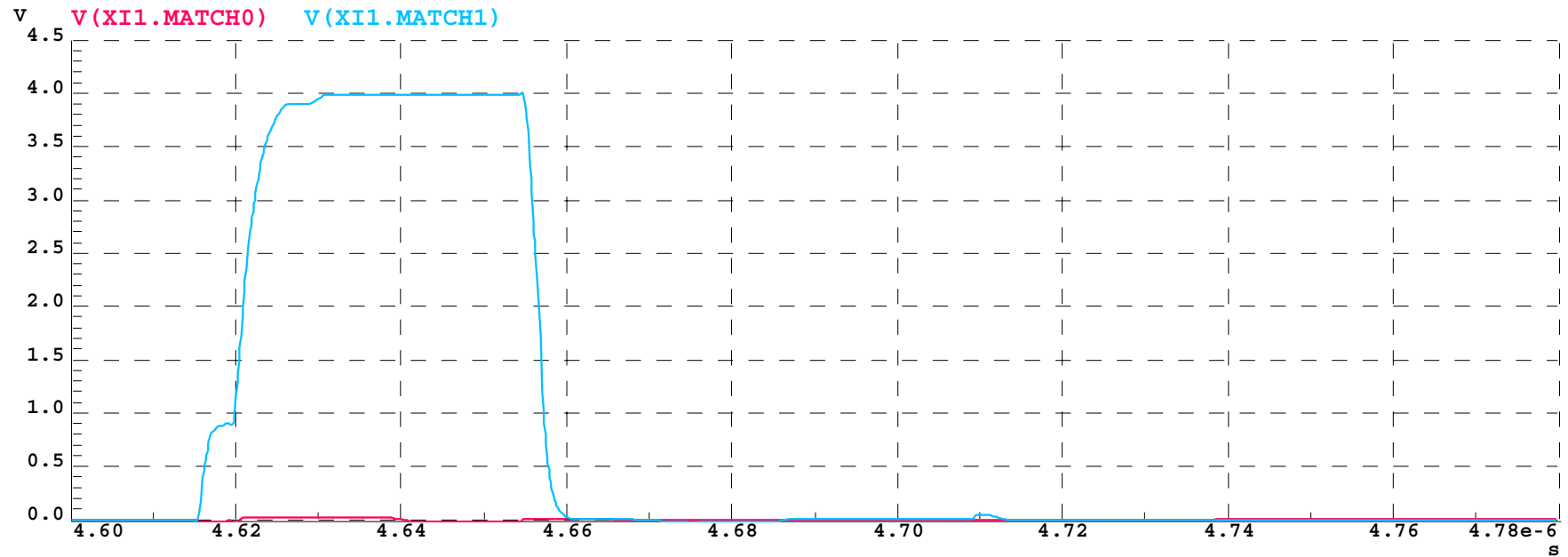
• Trigger match simulation of FE-D2D with XCK=40MHz, tt parameters, VDD=3V:



• Trigger match simulation of FE-D2D with XCK=40MHz, ss parameters, VDD=3V:



Trigger match simulation of FE-D2D with XCK=40MHz, iss parameters, VDD=4V:



Summary of 2D Scan Results for different corners

- For ff at 3V, see about 12ns in one direction, and 16ns in the other direction.
- For tt at 3V, see about 17ns in one direction, and 20ns in the other direction.
- For sf at 3V, see about 13ns in one direction, and 18ns in the other direction.
- For fs at 3V, see about 22ns in one direction, and 27ns in the other direction.
- For ss at 3V, see about 24ns in one direction, and 30ns in the other direction.
- For iss at 4V, see about 25ns in one direction, and 35ns in the other direction.

Conclusions:

- The large asymmetry in the two transition directions comes from the non-optimal sizing of the NAND and NOR transistors. However, this design seems to comfortably meet the requirements, and is the most compact (critical NMOS are min size, critical PMOS are twice min size).
- For the write scan, the 50ns constraint is very strict, otherwise the chip cannot operate reliably at 20MHz column clock frequency. For the read scan, the constraint is less strict, as it is really only the combined vertical and horizontal scan that must complete in less than 100ns (one ROCK period), and the horizontal scan is only allowed to occur during the last 50ns of this time.

Overall Conclusions

- Have studied most critical operations in the digital readout using large netlist, including all extracted parasitics (although not with SPICE simulation in analog).
- In FE-D2 design, basic design goals are met: successful 3V operation for all pre-rad corner models, and successful 4V operation for all post-rad corner models.
- There is very little margin left for operation beyond this point in FE-D2.

Faults:

- Weaknesses in the FE-D1 design have been studied, and fixed where possible.
- There was a FF in the CEU which was driving too large a load, and causing a glitch under worst case post-rad conditions. This was fixed by using Q and Qbar outputs for extra drive.
- There is non-optimal sizing of transistors in the sparse scan logic of the column pair and the EOC buffer. Here, one modification was made in FE-D2S, but it is not apparently optimal. However, this optimization is very sensitive to the relative degradation of the NMOS and PMOS in the worst case post-rad model, and so it is not an exact problem.
- There is an undersized NAND in EOC logic which is currently limiting the maximum operation speed of the design in simulation. There is not enough space in the present layout to increase the size of the NMOS in this gate.